Supervised Chemical Pattern Recognition in Almond (*Prunus dulcis*) Portuguese PDO Cultivars: a PCA and LDA Based Triennial Study

João C.M. Barreira[a,b], Susana Casal[b], Isabel C.F.R. Ferreira[a], António M. Peres[a,c], José Alberto Pereira[a,*], M. Beatriz P.P. Oliveira[b,*]

[a]*CIMO-ESAB, Instituto Politécnico de Bragança, Campus de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal.*

[b]*REQUIMTE, Departamento de Ciências Químicas, Faculdade de Farmácia da Universidade do Porto, Rua Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal.*

[c]*LSRE - Laboratory of Separation and Reaction Engineering - Associate Laboratory LSRE/LCM, Escola Superior Agrária, Instituto Politécnico de Bragança, Campus de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal.*

***Corresponding authors**

Telephone: +351 222078927 / +351 273303277

Fax: +351 222 003 977 / +351 273325405

E-mail: *beatoliv@ff.up.pt / jpereira@ipb.pt*

Abstract

Almonds harvested in three years in Trás-os-Montes (Portugal) were characterized to find differences among Protected Designation of Origin (PDO) *Amêndoa Douro* and commercial non-PDO cultivars. Nutritional parameters, fiber (neutral and acid detergent fibers, acid detergent lignin and cellulose), fatty acids, triacylglycerols (TAG) and tocopherols were evaluated. Fat was the major component, followed by carbohydrates, protein and moisture. Fatty acids were mostly detected as monounsaturated and polyunsaturated forms, with relevance of oleic and linoleic acids. Accordingly, 1,2,3-trioleoylglycerol and 1,2-dioleoyl-3-linoleoylglycerol were the major TAG. α-Tocopherol was the leading tocopherol. To verify statistical differences among PDO and non-PDO cultivars independently of the harvesting year, data were analyzed through an analysis of variance, a principal components analysis and a linear discriminant analysis (LDA). These differences identified classification parameters, providing an important tool for authenticity purposes. The best results were achieved with TAG analysis coupled with LDA, that proved its effectiveness to discriminate almond cultivars.

***Keywords**:* PDO almond, nutritional/chemical composition, authenticity, PCA, LDA.

INTRODUCTION

Almonds are the most widely consumed tree nuts. In Portugal, almond is an important product, with a production of 12 454 t spread through 38 444 ha, mainly located in "Terra Quente Transmontana" and Algarve.[1] Despite almonds' high fat content, 80% or more of the lipidic fraction is unsaturated, and the correspondent fatty acid profile might be cardio-protective. Nowadays, there are increasing experimental evidences suggesting that almonds improve serum lipid profiles and cholesterol status, reducing the risk of cardiovascular diseases.[2-4] Whereas the consumption of monounsaturated fatty acids (MUFA) decreases the risk of coronary diseases by 19%, the consumption of polyunsaturated fats decreases that risk by 38%.[4] Together with fatty acids profile and phytosterols[5] other bioactive compounds such as polyphenols[6-8] and tocopherols[7,8] may contribute to reduce the incidence of cardiovascular diseases[2,3] or reduce viral load in HIV-infected patients.[9]

Some studies have been conducted in American[10-14], Irish[7], Spanish[10,15,16], Italian[10,12,16], French[10,16], Australian[10] and Tunisian[10,12] cultivars, in which almond was characterized for having high amounts of fat (42-57%), protein (19-23%) and carbohydrates (20-27%), and low amounts of moisture (3-9%). Fiber and ash presented typical values of 11%[15] and 2.5-4.5%[10,13], respectively. Regarding fatty acids composition, almond presents mainly monounsaturated (~60%) and polyunsaturated (~30%) compounds.[14,15] However, the information related with nutritional and chemical characterization in Portuguese almond cultivars is still rather scarce. In fact, the available studies were dedicated to more specific features.[17-21] Hence, the chemical and nutritional compositions of selected regional almond cultivars of PDO *Amêndoa Douro* (*Casa Nova, Duro Italiano, Pegarinhos* one or two seeds and *Refego*) and commercial cultivars (*Ferraduel, Ferragnes, Ferrastar, Gloriette*

and *Marcona*) remain an interesting field of study, especially due to their high production levels and economic relevance.

Thus, the main objective of this work was the nutritional and chemical characterization of almond, allowing verifying chemical patterns that might act as fingerprints of *P. dulcis* PDO cultivars. The classification methods were based on the differences among chemical and/or nutritional contents among *Amêndoa Douro* (PDO) and commercial cultivars. To obtain a more comprehensive characterization, samples of three consecutive years were used, ensuring robustness against the influence of seasonal variability over nutritional and chemical parameters. So, chemical and nutritional data were tested using analysis of variance (ANOVA), principal components analysis (PCA), as a pattern recognition method and a stepwise linear discriminant analysis (LDA). The capability to authenticate almond cultivars is of great importance, either to conduct genetic improvement strategies, or to enhance their industrial applications and commercialization strategies.

MATERIALS AND METHODS

Standards and reagents

All reagents were of analytical grade purity: methanol and diethyl ether were supplied by Lab-Scan (Lisbon, Portugal); toluene from Riedel-de-Haen (Seelze, Germany); sulphuric acid from Fluka (Madrid, Spain). The fatty acids methyl ester (FAME) reference standard (47885-U) mixture (37 fatty acids C4 to C24) was from Supelco (Bellefonte, PA) and purchased from Sigma (St. Louis, MO), as also other individual fatty acid isomers.

Triacylglycerols 1,2,3-tripalmitoylglycerol (PPP), 1,2,3-tristearoylglycerol (SSS), 1,2,3-trilinolenoylglycerol (LnLnLn), and 1,2,3-tripalmitoleoylglycerol (PoPoPo), of purity

>98%, and 1,2,3-trioleoylglycerol (OOO), 1,2,3-trilinoleoylglycerol (LLL), 1,2-dilinoleoyl-3-palmitoyl-*rac*-glycerol (PLL), 1,2-dilinoleoyl-3-oleoyl-*rac*-glycerol (LLO), 1,2-dipalmitoyl-3-oleoyl-*rac*-glycerol (PPO), 1,2-dioleoyl-3-stearoyl-*rac*-glycerol (OOS), 1-palmitoyl-2-oleoyl-3-linoleoylglycerol (POL), and 1,2-dioleoyl-3-palmitoyl-*rac*-glycerol (POO), of ≈99% purity, were purchased from Sigma (St. Louis, MO). Acetonitrile and acetone were of HPLC grade and obtained from Merck (Darmstadt, Germany).

Tocopherols and tocotrienols (α, β, γ and δ) were purchase from Calbiochem (La Jolla, San Diego, CA). 2-Methyl-2-(4,8,12-trimethyltridecyl)chroman-6-ol (tocol) (Matreya Inc., Pleasant Gap, PA) was used as internal standard (IS). Butylated hydroxytoluene (BHT) was obtained from Aldrich (Madrid, Spain), hexane was of HPLC grade from Merck (Darmstad, Germany) and 1,4-dioxane was from Fluka (Madrid, Spain). All the other chemicals were of analytical grade and obtained from Sigma Chemical Co. (St. Louis, MO). Water was treated in a Mili-Q water purification system (TGI Pure Water Systems, Brea, CA).

Samples and sample preparation

Almonds were obtained from selected PDO (*Casa Nova, Duro Italiano, Pegarinhos* one or two seeds and *Refego*) and commercial (*Ferraduel, Ferragnes, Ferrastar, Gloriette* and *Marcona*) cultivars and collected in August-September during three years (2006, 2007 and 2008) in orchards located in Southwest Trás-os-Montes, Northeast Portugal. For each cultivar 50 almonds were collected and divided in 2 groups. Samples of each cultivar were obtained from five selected trees (the same trees were selected over the three years, except for *Refego*, *Gloriette* and *Marcona*, which were not available in 2006). Selected plants were not irrigated and no phytosanitary treatments were applied. The fruits were dried at room temperature and exposed to sun, in accordance with the traditional and common

practices in the region. Almonds were kept at -20 ºC and protected from light until further use. Immediately before analysis, almonds were chopped to obtain a fine dried powder (20 mesh).

Proximate analysis

The chemical composition (moisture, protein, fat, ash, fiber) of almonds was determined using the AOAC procedures.[22] The crude protein content of the samples was estimated by the macro Kjeldahl method; the crude fat was determined by extracting a known weight of powdered almond sample with petroleum ether (bp 40-60 °C), using an Universal extraction system B-811 (Büchi, Flawil, Switzerland); the ash content was determined by incineration at 550±15 ºC until whitish ash appear. Neutral detergent fiber (NDF), including cellulose, hemicelluloses and lignin, acid detergent fiber (ADF), including cellulose and lignin less digestible and woody fibers and acid detergent lignin (ADL) were determined according to Robertson and Van Soest method[23] with minor changes. Total carbohydrates were calculated by difference: Total carbohydrates = 100 – (g moisture + g protein + g fat + g ash + g of fiber). Total energy was calculated according to the following equation: Energy (kcal) = 4 × (g protein + g carbohydrate) + 9 × (g lipid).[24]

Oil extraction procedure

Almonds were manually shelled and then chopped in a 643 MX coffee mill (Moulinex, Spain). Crude oil was obtained from finely chopped almonds (≈5 g, with anhydrous sodium sulfate), extracted with light petroleum ether (bp 40-60 °C) during 1.5 h (for the determination of total fat content the extraction time was 24 h) in a Universal extraction system B-811 (Büchi, Switzerland); the residual solvent was removed by flushing with

nitrogen. This oil was used for the evaluation of fatty acids, triacylglycerols and tocopherol contents, as follows.

Fatty acids analysis

Fatty acid methyl esters (FAMEs) were prepared by oil hydrolysis with a 2 M methanolic potassium hydroxide solution, and extraction with n-heptane, in accordance with ISO 5509 method[25] and following a procedure previously described by the authors.[26] The fatty acid profile was analyzed with a Chrompack CP 9001 chromatograph (Chrompack, Middelburg, Netherlands) equipped with a split-splitless injector, a flame ionization detector (FID), and a Chrompack CP-9050 autosampler. The results are expressed in relative percentage of each fatty acid, calculated by internal normalization of the chromatographic peak area, and assuming that the detector response was the same for all compounds.

Triacylglycerol analysis

The chromatographic analyses were performed according to the procedure previously described[26], with a Jasco (Tokyo, Japan) HPLC system, equipped with a PU-1580 quaternary pump and a Jasco AS-950 automatic sampler with a 10 μL loop. The chromatographic separation of the compounds was achieved with a Kromasil 100 C_{18} (5 μm; 250×4.6 mm) column (Teknokroma, Barcelona, Spain) operating at room temperature (≈20 °C). Detection was performed with an evaporative light-scattering detector (ELSD) (model 75-Sedere, Alfortville, France). Taking into account the selectivities (R, relative retention times to LLL), peaks were identified according to the logarithms of R in relation to homogeneous TAG standards. Quantification of the peaks was made by internal

normalization of chromatographic peak area, and the results are expressed in relative percentage, assuming that the detector response was the same for all compounds.

Tocopherols analysis

An oil solution in hexane with an adequate amount of internal standard was prepared and analyzed by HPLC in a normal-phase column (Inertsil 5 SI, 250×3 mm) from Varian (Middelburg, The Netherlands) operating at room temperature. The HPLC equipment consisted of an integrated system with a PU-980 pump, an AS-950 auto-sampler, an MD-910 multiwavelenght diode array detector (DAD) connected in series with an FP-920 fluorescence detector (Jasco, Japan) programmed for excitation at 290 nm and emission at 330 nm, gain 10. Data were analyzed using Borwin-PDA Controller Software (JMBS, France). The chromatographic separation was achieved following the procedure previously described[27]. The compounds were identified by chromatographic comparisons with authentic standards and by their UV spectra. Quantification was based on the fluorescence signal response, using the internal standard method.

Statistical analysis

All analyses (extractions) were performed in duplicate; each replicate was quantified also in duplicate (samples for *Gloriette*, *Marcona* and *Refego* were not available in 2006). Data were expressed as means±standard deviations. All statistical tests were performed at a 5% significance level using the SPSS software, version 18.0 (SPSS Inc).

The fulfillment of the one-way ANOVA requirements, specifically the normal distribution of the residuals and the homogeneity of variance, was tested by means of the Kolmogorov-Smirnov with Lilliefors correction or the Shapiro-Wilk's (depending on the amount of samples), and the Levene's tests, respectively. In the cases where statistical significance

differences were identified, the dependent variable were compared using Tukey's honestly significant difference (HSD) or Tamhane's T2 multiple comparison tests, when homoscedasticity was verified or not, respectively.

PCA was applied as pattern recognition unsupervised classification method. PCA transforms the original, measured variables into new uncorrelated variables called principal components. The first principal component covers as much of the variation in the data as possible. The second principal component is orthogonal to the first and covers as much of the remaining variation as possible, and so on.[28] The number of dimensions to keep for data analysis was evaluated by the respective eigenvalues (which should be greater than one), by the Cronbach's alpha parameter (that must be positive) and also by the total percentage of variance (that should be as higher as possible) explained by the number of components selected.

LDA was used as a supervised learning technique to classify *P. dulcis* cultivars according to their nutritional, fatty acids, triacylglycerols or tocopherols contents. The assumptions of LDA, which include linear relationship between all pairs of independent variables, the normality within groups and homogeneity of variances and of variance–covariance matrices, were checked using the Kolmogorov-Smirnov with Lilliefors correction, the Levene and M-Box tests, respectively.[29] A stepwise technique, using the Wilks' λ method with the usual probabilities of F (3.84 to enter and 2.71 to remove), was applied for variable selection. This procedure uses a combination of forward selection and backward elimination procedures, where before selecting a new variable to be included, it is verified whether all variables previously selected remain significant.[30-32] Discriminant analysis defines a combination of variables in a way that the first function furnishes the most general discrimination between groups, the second provides the second most, and so on.[33] With this approach, it is possible to identify the significant variables among the nutritional,

fatty acids, triacylglycerols and tocopherols profiles obtained for each sample. To verify which canonical discriminant functions were significant, the Wilks' λ test was applied. To avoid overoptimistic data modulation, a leave-one-out cross-validation procedure was carried out to assess the model performance. Moreover, the sensitivity and specificity of the discriminant model were computed from the number of individuals correctly predicted as belonging to an assigned group[30,32]. Sensitivity and specificity were calculated as follows:

$$\text{Sensitivity} = \frac{\text{number of samples of a specific group correctly classified}}{\text{total number of samples belonging to that specific group}}$$

$$\text{Specificity} = \frac{\text{number of samples of a specific group classified as belonging to this group}}{\text{total number of samples of any group classified as belonging to that group}}$$

RESULTS AND DISCUSSION

Proximate analysis

Table 1 shows the triennial means obtained for proximate composition of PDO cultivars and each single commercial cultivar. In general, fat is clearly the major component, crude protein and carbohydrates are present in similar contents (~20%), water, ash and fiber laid under 5%, conducting to high energy values (more than 610 kcal/100 g of fresh fruit). ADL was also detected but in minute amounts (<0.1 g/100 g of fresh fruit). This compositional profile is in agreement with previous results.[11,13,14,16]

The results from the one-way ANOVA analysis showed that, at a significance level of 5%, there were no differences (except for ash content) between the mean values of the chemical composition between PDO and commercial cultivars under study. In fact, no particular

tendency could be observed for the evaluated parameters. In the particular case of ash content, Tamhane's T2 test indicated that the tested samples were classified equally. These results seem to indicate that proximate composition data possessed very limited differentiation ability regarding almond cultivar discrimination.

Fatty acids analysis

Table 2 shows the triennial means obtained for fatty acids profiles of each commercial cultivar and for the PDO cultivars. Besides the fatty acids reported in **Table 2**, C14:0, C15:0, C17:0, C20:1, C18:3, C21:0, C22:0, C20:3 and C24:0, were also detected but only in trace amounts (< 0.1%). These results showed that almond fat is mainly constituted by three fatty acids: oleic (C18:1), linoleic (C18:2) and palmitic (C16:0) acids accounting for more than 96% of the total FA content, a value analogous to those obtained by other research groups.[3,10-13,16] The analysis carried out showed that the residuals followed a normal distribution (P>0.05) and, except for three fatty acids (C17:1, C18:0 and C20:0), there was heteroscedasticity. For some fatty acids, one-way ANOVA allowed finding evidences of significant statistical differences between their contents in PDO cultivars and those of commercial cultivars (P<0.05). Based on the results from the Tamhanes' T2 test (P<0.05) it was found that *Ferrastar* and *Gloriette* had the lowest C16:1 and C18:0 levels, respectively; PDO cultivars presented the lowest C18:1 and the highest C18:2 contents; *Gloriette* has significantly less SFA than PDO and *Ferrastar* cultivars; PDO cultivars presented lower MUFA and higher PUFA than *Ferraduel* and *Gloriette* cultivars.

The low number of statistically significant differences among fatty acids profiles in PDO and non-PDO cultivars indicate that this parameter should be unsuitable for almond cultivars discrimination.

Triacylglycerols analysis

Table 3 shows the triennial means obtained for triacylglycerols (TAG) profiles of each commercial cultivar and for the PDO cultivars. The analysis carried out showed that the residuals had a normal distribution (P>0.05) for OLO, OOO and OOP, and, except for OOP and POP, the Levene test showed the assumption of equality of variances could not be assumed. Even so, to uniformize the analysis, when statistical signifcant differences (P<0.05) were detected by the one-way ANOVA test, differences among individual cultivars (PDO and commercial) were tested by means of the Tamhanes' T2 test instead of using the Tukey's test. The multiple comparisons test allowed obtaining general conclusions for almost all cases evaluated: PDO has the highest OLL and LLP contents; OLO presented the lowest value in *Ferraduel*, while LOP showed minimal values in *Gloriette* and *Marcona*. PLP reached maximal contents in *Marcona*, whilst OOO presented its lowest value in PDO cultivars. SOO was higher in *Ferrastar* than in all the remaining cultivars.

Further than the previous considerations, the results confirmed the prevalence of OOO and OLO. In general, the detected profiles are comparable to previous publications[34,35], and are in accordance with the previously described FA composition.

The observed differences indicate that TAG profiles may be useful as a practical classification tool for almond cultivars discrimination, namely between PDO and commercial cultivars as well among the last ones (see *3.5. PCA and LDA*).

Tocopherols and tocotrienols analysis

Table 4 shows the triennial means obtained for triacylglycerols and tocotrienols profiles of each commercial cultivar and PDO cultivars. The mean values of all vitamers (except α-tocotrienol) did not show significant differences among the assayed cultivars. Concerning

the results of the Levene test, for α-tocotrienol, γ-tocopherol and γ-tocotrienol, the data showed heteroscedasticity and so, the few significant statistical differences detected by means of the one-way ANOVA (P<0.05) were evaluated based on the Tamhanes' T2 test. Globally at a 5% significance level and from a statistical point of view, α-tocotrienol content was greater in PDO cultivars, which is in agreement with previous results showing that the effect of the specific characteristics of the genotype might affect the amounts of each tocopherol homologue.[36]

In general, α-tocopherol was the major compound followed by γ-tocopherol. On the other hand, δ-tocopherol was the minor vitamer in all cultivars. However, the obtained results did not reveal potential to discriminate PDO and commercial cultivars. The results obtained for the triennial averages are comparable to previously published works.[11,14,36]

Overall and independently of the harvesting year, almonds have high caloric values, over 610 kcal/100 g of fresh weight, providing a powerful energy source. The FA profiles were similar for commercial and PDO cultivars, with *oleic* (C18:1ω9), *linoleic* (C18:2 ω6) and palmitic acid (C16:0) as the compounds present in major amounts. Fatty acid profiles were reflected in TAG composition, with OOO, OLO and OLL as predominant compounds.

In general, the results highlight almond as a promising source of bioactive compounds, improving its commercial value.

PCA and LDA

The previous analysis showed that among the evaluated parameters (proximate analysis data, fatty acids profile, triacylglycerols, tocopherols and tocotrienols analysis), the TAG data recorded for the PDO and non-PDO cultivars possessed the higher discrimination

potential. Therefore, it was decided to only use these data for evaluating both unsupervised and supervised classification techniques, namely PCA and LDA.

The number of dimensions considered for PCA was chosen in order to keep it small enough so that meaningful interpretations were possible, and by ensuring their reliability, assessed by the value of the Cronbach's alpha parameter as well as by the related eigenvalue. The biplot of component loadings (**figure 1**) indicate that the first two dimensions account for most of the variance of all quantified variables (44.7% and 26.3%, respectively). The selection of only two dimensions was supported in the observation that for higher dimensions negative Cronbach's alpha values (-0.089, for the third dimension) and eigenvalues lower than one (0.926, for the third dimension) were obtained (data not shown). The first dimension is positively associated with OLO, LLP, LLL, OLL and PLP. So, as can be seen from **figure 1**, these variables have a high impact especially within the PDO cultivars. On the other hand, OOO and OOP are very negatively scored for the first dimension, showing a significative impact especially for non PDO cultivars, namely *Ferraduel*, *Gloriette* and *Marcona*. The second dimension is mostly related with the quantified variables LOP and SOO in the positive region and POP in the negative region. In accordance, SOO and LOP highly accounted for non PDO cultivars (*e.g.*, *Ferrastar* and *Ferragnes*) and POP accounted for PDO cultivars.

Regarding the relation between the objects and variables (**figure 1**), it is clear that *Ferrastar*, *Ferraduel*, *Gloriette* are characterized for having, respectively, high SOO, OOO and OOP contents, while PDO present the highest levels of LLL and OLL.

Although, the lower dimensional solutions often conceal differences among variables, the PCA results were satisfactory, and there was no need to increase the number of dimensions. In fact, the results plotted in **figure 1** show that, in general, the TAG profiles recorded for the PDO and non PDO cultivars evaluated in this study possess valuable

information that may be used as an effective tool for diferentiating samples of almonds from PDO cultivars (black lines in **figure 1**) from those of non-PDO cultivars (grey dot and dash lines in **figure 1**).

A LDA was also performed to infer about which chemical and nutricional parameters evaluated possessed discriminative ability that would allow differentiating PDO/non-PDO cultivars. Before the analysis, the fulfillment of the LDA assumptions was checked. Nevertheless it should be noticed that, although this method requires the normality of the data, it can deal with deviations from normality, having good robustness.

The significant independent variables (parameters) were selected using the stepwise procedure of the LDA, according to the Wilks' λ test. Only those that showed a statistical significant classification performance (P<0.05) were kept for analysis. Therefore, the LDA was carried out considering different combinations of the assayed parameters, in order to find which one discriminates better *Amêndoa Douro* (PDO cultivars) and commercial cultivars. The analysis showed that only TAG were used for the final discriminant model, being kept 7 of the 10 parameters evaluated (LLL, OLL and OOP, were not used). The model had only three significant discriminant functions (P<0.001 for the Wilks' λ test), which explained 97.8% of the total variance of the experimental data (the first explained 55.0%, the second 25.6% and the third 17.3%) (**Figure 2**).

The first function separates clearly *Ferrastar* cultivar (means of the canonical variance (MCV): PDO = 0.733; *Ferraduel* = -4.432; *Ferragnes* = -0.258; *Ferrastar* = 5.293; *Gloriette* = -3.643; *Marcona* = -2.391), and revealed to be more powerfully correlated with SOO. The second function separates mainly *Ferragnes* and *Gloriette* from the other cultivars (MCV: PDO = -0.093; *Ferraduel* = 1.157; *Ferragnes* = -4.346; *Ferrastar* = 1.676; *Gloriette* = -0.130; *Marcona* = 3.049) and showed to be more correlated with LOP. The third function separates acceptably PDO (MCV: PDO = -1.256; *Ferraduel* = 0.348;

Ferragnes = 1.169; *Ferrastar* = 2.446; *Gloriette* = 2.715; *Marcona* = 0.133), showing higher correlation with SOO and PLP.

In summary, as shown in **Figure 2**, samples belonging to PDO cultivars are all assembled within a single group quite apart from the other commercial cultivars. Indeed, the model showed a good classification performance, allowing to correctly classify (sensitivity) 94.4% of the samples within the leaving-one-out cross-validation procedure, regardless the harvesting year (**Table 5**).

In fact, as can be observed from the results reported in **Table 5**, only 3 of the 28 almond samples from PDO cultivars were misclassified: one classified as *Ferragnes* cultivar, another as *Marcona* cultivar and the latter as *Ferraduel* cultivar (with group probabilities equal to 0.905, 0.929 and 0.793, respectively). However, the results obtained for these same misclassified PDO almond samples also showed that the alternative classification group would be the right one (PDO group), although with lower group probabilities than the misclassification (0.091, 0.067 and 0.189, respectively). Therefore, these misclassifications were attributed to analytical errors, since the data obtained from repeated analysis of the sample picked in the same year allowed a correct classification. Furthermore, it should be remarked that no commercial cultivar sample was misclassified as other commercial or PDO cultivar, which reinforces the idea that TAG profile may be used as a practical tool for ensuring PDO samples authenticity. Finally, the satisfactory performance of the proposed classification procedure is also confirmed by the high overall specificity achieved (92%) for the cross-validation procedure.

Overall, it appears that genetically defined features may overcome the climatic conditions effect, probably because the assembly of all regional (PDO) cultivars resulted in higher variability among the values obtained for the assayed parameters. The higher broadness

defined for each parameter hindered the main purpose of obtaining a distinctive chemical pattern (independently of the possible effect of the harvesting year) with the ability to separate PDO and individual commercial cultivars. However, there are some distinctive features, mainly associated with TAG profiles. In fact, TAG contents allowed establishing a satisfactory classification model of almond cultivars as PDO or belonging to a specific commercial cultivar. The results showed that the discrimination model proposed can be used as a tool for differentiating PDO *Amêndoa Douro* cultivars from commercial almond cultivars. Nevertheless, since no external validation was carried out, the developed model should be used with some precaution. So, it was shown that almond consumers, producers or even food industry that uses almonds may use the proposed approach to prevent possible frauds, avoiding buying/selling lower-valuable commercially almonds as PDO almonds. This finding even more advantageous since almond cultivars may be correctly classified performing a single, fast and reliable assay (TAG analysis coupled with LDA).

Furthermore, this work represents a contribution to almond chemical and nutritional characterization. The obtained data may be useful in updating databases and composition tables. The complete characterization of almond cultivars represents important benefits, either from the correct diet definitions perspective, but also in the improvement of the technological processes and industrial applications.

REFERENCES

(1) Agricultural Statistics. *Instituto Nacional de Estatística I. P.* **2010**. Lisbon, Portugal.

(2) Albert, C. M.; Gaziano, J. M.; Willett, W. C.; Manson, J. E. Nut consumption and decreased risk of sudden cardiac death in the Physicians' Health Study. *Arch. Intern. Med.* **2002**, *162*, 1382-1387.

(3) Kris-Etherton, P. M.; Zhao, G.; Binkoski, A. E., Coval, S. M.; Etherton, T. D. The effect of nuts on coronary heart disease risk. *Nutr. Rev.* **2001**, *59*, 103-111.

(4) Rajaram, S.; Burke, K.; Connell, B.; Myint, T.; Sabate, J. A monounsaturated fatty acid-rich pecan-enriched diet favourably alters the serum lipid profile of healthy men and women. *J. Nutr.* **2001,** *131*, 2275-2279.

(5) Bruckert, E. Les phytostérols, place dans la prise en charge du patient hyperlipidémique. Aliments fonctionnels et lipides. *O.C.L.* **2001**, *8*, 312-316.

(6) Barreira, J. C. M.; Ferreira, I. C. F. R.; Oliveira, M. B. P. P.; Pereira, J. A. Antioxidant activity and bioactive compounds of ten Portuguese regional and commercial almond cultivars. *Food Chem. Toxicol.* **2008**, *46*, 2230-2235.

(7) Maguire, L. S.; O'Sullivan, S. M.; Galvin, K.; O'Connor, T. P.; O'Brien, N. M. Fatty acid profile, tocopherol, squalene and phytosterol content of walnuts, almonds, peanuts, hazelnuts and the macadamia nut. *Int. J. Food Sci. Nutr.* **2004**, *55*, 171-178.

(8) Kormsteiner, M.; Wagner, K. H.; Elmadfa, I. Tocopherols and total phenolics in 10 different nut types. *Food Chem.* **2006**, *98*, 381-387.

(9) Allard, J. P.; Aghdassi, E.; Chau, J.; Tam, C.; Kovacs, C. M.; Salit, I. E.; Walmsley, S. L. Effects of vitamin E and C supplementation on oxidative stress and viral load in HIV-infected subjects. *AIDS* **1998**, *12*, 1653-1659.

(10) Ahrens, S.; Venkatachalam, M.; Mistry, A. M.; Lapsley, K.; Sathe, S. K. Almond (*Prunus dulcis* L.) protein quality. *Plant Food Hum. Nutr.* **2005**, *60*, 123-128.

(11) López-Ortiz, C. M., Prats-Moya, S., Sanahuja, A. B., Maestre-Pérez, S. E., Grané-Teruel, N.; Martín-Carratalá, M. L. Comparative study of tocopherol homologue content in four almond oil cultivars during two consecutive years. *J. Food Compos. Anal.* **2008**, *21*, 144-151.

(12) Martín-Carratalá, M. L.; García-López, C.; Berenguer-Navarro, V.; Grané-Teruel, N. New Contribution to the Chemometric Characterization of Almond Cultivars on the Basis of Their Fatty Acid Profiles. *J. Agric. Food Chem.* **1998**, *46*, 963-967.

(13) Venkatachalam, A.; Sathe, S. K. Chemical Composition of Selected Edible Nut Seeds. *J. Agric. Food Chem.* **2006**, *54*, 4705-4714.

(14) Chen, C.-Y.; Lapsley, K.; Blumberg, J. A nutrition and health perspective on almonds. *J. Sci. Food Agr.* **2006**, *86*, 2245-2250.

(15) Cherif, A.; Belkacemi, K.; Kallel, H.; Angers, P.; Arul, J.; Boukhchina, S. Phytosterols, unsaturated fatty acid composition and accumulation in the almond kernel during harvesting period: Importance for development regulation. *C. R. Biol.* **2009**, *332*, 1069-1077.

(16) Piscopo, A.; Romeo, F. V.; Petrovicova, B.; Poiana, M. Effect of the harvest time on kernel quality of several almond varieties (*Prunus dulcis* (Mill.) D.A. Webb). *Sci. Hortic.* **2010**, *125*, 41-46.

(17) Cordeiro, V.; Monteiro, A. Almond growing in Trás-os-Montes region (Portugal). *Acta Hortic.* **2001**, *591*, 161-165.

(18) Dourado, F.; Barros, A.; Mota, M.; Coimbra, M. A.; Gama, F. M. Anatomy and Cell Wall Polysaccharides of Almond (*Prunus dulcis* D. A. Webb) Seeds. *J. Agr. Food Chem.* **2004a**, *52,* 1364-1370.

(19) Dourado, F.; Madureira, P.; Carvalho, V.; Coelho, R.; Coimbra, M. A.; Vilanova, M.; Mota, M.; Gama, F. M. Purification, structure and immunobiological activity of an arabinan-rich pectic polysaccharide from the cell walls of *Prunus dulcis* seeds. *Carbohyd. Res.* **2004a**, *339,* 2555-2566.

(20) Estevinho, B. N.; Ratola, N.; Alves, A.; Santos, L. Pentachlorophenol removal from aqueous matrices by sorption with almond shell residues. *J. Hazard. Mater.* **2006**, *137*, 1175-1181.

(21) Martins, M.; Tenreiro, R.; Oliveira, M. M. Genetic relatedness of Portuguese almond cultivars assessed by RAPD and ISSR markers. *Plant Cell Rep.* **2003**, *1*, 71-78.

(22) AOAC. *Official methods of analysis* (16th Ed.), **1995**. Arlington VA, USA: Association of Official Analytical Chemists.

(23) Robertson, J. B.; Van Soest, P. J. The detergent system analysis and its application to human foods. In: James, W. P. T., Theander, O. (eds.). The Analysis of Dietary Fiber in Food. Vol. 3, Marcel Dekker, Inc., **1981**, New York, pp. 123-158.

(24) Portuguese Government Legislation: *Decreto-Lei* 167, **2004**, July 17th. *Ministério da Agricultura, Desenvolvimento Rural e Pescas*. *Diário da República nº158*, 4080-4083.

(25) ISO, International Organization for Standardization 5509, **2000** - Animal and vegetable fats and oils - Preparation of methyl esters of fatty acids.

(26) Barreira, J. C. M.; Casal, S.; Ferreira, I. C. F. R.; Oliveira, M. B. P. P.; Pereira, J. A. Nutritional, fatty acid and triacylglycerol profiles of *Castanea sativa* Mill. cultivars: a compositional and chemometric approach. *J. Agric. Food Chem.* **2009a**, *57*, 2836-2842.

(27) Barreira, J. C. M.; Alves, R. C.; Casal, S.; Ferreira, I. C. F. R.; Oliveira, M. B. P. P.; Pereira, J. A. Vitamin E profile as a reliable authenticity discrimination factor between chestnut (*Castanea sativa* Mill.) cultivars. *J. Agric. Food Chem.* **2009b**, *57*, 5524-5528.

(28) Patras, A.; Brunton, N. P.; Downey, G.; Rawson, A.; Warriner, K.; Gernigon, G. Application of principal component and hierarchical cluster analysis to classify fruits

and vegetables commonly consumed in Ireland based on in vitro antioxidant activity. *J. Food Compos. Anal.* **2011**, *24*, 250-256.

(29) Hill, T.; Lewicki, P. Statistics: methods and applications. A comprehensive reference for science, industry, and data mining, StatSoft, **2006**.

(30) López, A.; García, P.; Garrido, A. Multivariate characterization of table olives according to their mineral nutrient composition. *Food Chem.*, **2008**, *106*, 369-378.

(31) Maroco, J. Análise Estatística, com utilização do SPSS, Edições Sílabo, **2003**, Lisboa, Portugal.

(32) Rencher, A. C. Methods of Multivariate Analysis, John Willey, **1995**, New York.

(33) Benitez, E.; Nogales, R.; Campos, M.; Ruano, F. Biochemical variability of olive-orchard soils under different management systems. *Appl. Soil Ecol.* **2006**, *32*, 221-231.

(34) Prats-Moya, M. S.; Grané-Teruel, N.; Berenguer-Navarro, V.; Martín-Carratalá, M. L. A chemometric study of genotypic variation in triacylglycerol composition among selected almond cultivars. *J. Am. Oil Chem. Soc.* **1999**, *76*, 267-272.

(35) Martín-Carratalá, M. L.; Llorens-Jordá, C.; Berenguer-Navarro, V.; Grané-Teruel, N. Comparative study on the triglyceride composition of almond kernel oil. A new basis for cultivar chemometric characterization. *J. Agric. Food Chem.* **1999**, *47,* 3688-3692.

(36) Kodad, O.; Estopañán, G.; Juan, T.; Mamouni, A.; Company, R.S. Tocopherol Concentration in almond oil: genetic variation and environmental effects under warm conditions. *J. Agric. Food Chem.* **2011**, *59*, 6137-6141.

The authors are grateful to FCT, POPH-QREN and FSE for financial support to J.C.M. Barreira (SFRH/BPD/72802/2010).

Figure 1. Projections of the Average Scores of Almond Cultivars for the Two Rotated Principal Components. Objects and Component Loadings were Biplotted Using Sample Origin as Labeling Variable. Frd- *Ferraduel*; Frg- *Ferragnes*; Frs- *Ferrastar*; Glt- *Gloriette*; Mrc- *Marcona*.

Figure 2. Mean Scores of Almond Cultivars Projected for the Three Rotated Discriminant Functions Defined from TAG Profiles. Frd- *Ferraduel*; Frg- *Ferragnes*; Frs- *Ferrastar*; Glt- *Gloriette*; Mrc- *Marcona*.

Table 1.

Proximate Composition (g/100 g fresh weight) and Corresponding Energy (*per* 100 g fresh weight). The results are presented as mean±SD.

		Water	Fat	Protein	Carbohydrates	NDF	ADF	Cellulose	Ash[a]	Energy (kcal)
Cultivar	PDO (n = 28)	5±1	50±6	23±2	20±5	2.9±0.5	0.5±0.2	0.4±0.2	3.1±0.2 a	618±31
	Ferraduel (n = 6)	4±1	52±3	22±4	20±2	3±1	0.5±0.2	0.4±0.2	2.9±0.2 a	633±11
	Ferragnes (n = 6)	4±1	50±7	21±2	21±6	2.8±0.4	0.4±0.1	0.4±0.1	2.8±0.3 a	622±43
	Ferrastar (n = 6)	4±1	51±2	23±4	18±3	3±1	0.4±0.1	0.4±0.1	2.9±0.2 a	626±17
	Gloriette (n = 4)	4.5±0.5	49±4	23±1	20±4	3.2±0.2	0.4±0.2	0.4±0.2	3.2±0.2 a	615±21
	Marcona (n = 4)	4±1	55±2	24±2	14±2	2.6±0.5	0.4±0.2	0.4±0.2	2.8±0.1 a	647±13
Homoscedasticity[b]	*P*-value	0.051	0.017	0.016	0.280	0.035	0.444	0.331	0.269	0.013
Normal distribution[c]	*P*-value	0.200[e]	0.055	0.192	0.200[e]	0.029	0.007	0.003	0.200[e]	0.002
One-way ANOVA[d]	*P*-value	0.698	0.475	0.621	0.202	0.824	0.987	0.984	0.025	0.456

[a]Means were evaluated using the Levene multiple comparison test.
[b]Homoscedasticity among cultivars was tested by means of the Levene test: homoscedasticity, *P*-value>0.05; heteroscedasticity, *P*-value<0.05.
[c]Normal distribution of the residuals was evaluated using Kolmogorov-Smirnov with Lilliefors correction test (n>20).
[d]*P*<0.05 meaning that the mean value of the evaluated parameter of at least one cultivar differs from the others (in this case multiple comparison tests were performed).
[e]This is a lower bound of the true significance.

Table 2.

Fatty Acids Composition for Assembled PDO and Individual non-PDO Cultivars (%). The results, analysed through 1-way ANOVA, are presented as mean±SD.[a]

		C16:0	C16:1	C17:1	C18:0	C18:1	C18:2	C20:0	SFA	MUFA	PUFA
Cultivar	PDO (n = 28)	6.9±0.5 a	0.5±0.1 ab	0.11±0.01	2.3±0.4 a	65±5 b	25±4 a	0.11±0.03	9.5±0.5 a	66±5 b	25±4 a
	Ferraduel (n = 6)	6.5±0.5 ab	0.5±0.1 ab	0.11±0.01	2.0±0.3 ab	71±3 a	20±2 b	0.11±0.03	9±1 ab	71±3 a	20±2 b
	Ferragnes (n = 6)	7±1 ab	0.5±0.1 ab	0.10±0.01	2.3±0.3 a	68±6 ab	22±5 ab	0.12±0.02	9±1 ab	69±6 ab	22±5 ab
	Ferrastar (n = 6)	6.6±0.4 ab	0.38±0.03 b	0.11±0.01	2.3±0.3 a	68±4 ab	22±4 ab	0.11±0.01	9.1±0.4 a	69±4 ab	22±4 ab
	Gloriette (n = 4)	6.0±0.1 ab	0.46±0.01 ab	0.11±0.01	1.5±0.1 b	74±2 a	17.6±0.3 b	0.09±0.01	7.7±0.1 b	75±1 a	18±1 b
	Marcona (n = 4)	6.8±0.2 a	0.59±0.05 a	0.11±0.01	2.0±0.2 ab	69±2 ab	22±2 ab	0.09±0.01	9.0±0.5 ab	69±2 ab	22±2 ab
Homoscedasticity[b]	*P*-value	0.046	0.029	0.938	0.582	0.017	0.015	0.053	0.011	0.019	0.015
Normal distribution[c]	*P*-value	0.200[e]	0.200[e]	0.080	0.200[e]	0.200[e]	0.094	0.200[e]	0.200[e]	0.200[e]	0.097
One-way ANOVA[d]	*P*-value	0.005	0.011	0.354	0.001	0.003	0.013	0.498	<0.001	0.003	0.013

[a]Means within a column with different letters differ significantly ($p < 0.05$), evaluated either using the multiple comparison Tukey's HSD or Tamhane's T2 tests, depending on the fulfilment or not of the homoscedasticity requirement, respectively.
[b]Homoscedasticity among cultivars was tested by means of the Levene test: homoscedasticity, *P*-value>0.05; heteroscedasticity, *P*-value<0.05.
[c]Normal distribution of the residuals was evaluated using Kolmogorov-Smirnov with Lilliefors correction test (n>20).
[d]*P*<0.05 meaning that the mean value of the evaluated parameter of at least one cultivar differs from the others (in this case multiple comparison tests were performed).
[e]This is a lower bound of the true significance.

Table 3.

Triacylglycerol Composition (%) for Assembled PDO and Individual non-PDO Cultivars. The results, analysed through 1-way ANOVA, are presented as mean±SD.[a]

		LLL	OLL	LLP	OLO	LOP	PLP	OOO	OOP	POP	SOO
Cultivar	PDO (n = 28)	0.9±0.3	15±5 a	0.7±0.3 a	29±4 a	5±2 bc	2±1 ab	38±8 d	7±2	0.4±0.2	0.74±0.05 b
	Ferraduel (n = 6)	0.37±0.05	7±1 c	0.4±0.1 c	25±1 b	4±1 cd	0.06±0.01 c	55±3 a	7±1	0.029±0.001	0.7±0.1 b
	Ferragnes (n = 6)	0.79±0.05	14±2 ab	0.72±0.04 ab	32±2 a	8±1 a	0.06±0.01 c	39±3 cd	6±1	0.08±0.01	0.83±0.05 b
	Ferrastar (n = 6)	0.5±0.2	10.3±0.5 b	0.4±0.1 c	33±2 a	7±1 a	0.08±0.01 c	41±1 cd	5±1	0.07±0.01	2.2±0.3 a
	Gloriette (n = 4)	0.43±0.02	7±1 c	0.39±0.05 c	30±1 a	3.4±0.4 d	0.22±0.05 c	52±1 ab	6±1	0.46±0.05	0.7±0.3 b
	Marcona (n = 4)	0.5±0.2	10±3 abc	0.5±0.2 abc	28±1 a	2.8±0.4 d	3.2±0.5 a	47±3 abc	7±1	0.21±0.05	0.4±0.1 b
Homoscedasticity[b]	*P*-value	<0.001	<0.001	<0.001	0.001	0.015	<0.001	0.002	0.132	0.052	<0.001
Normal distribution[c]	*P*-value	0.001	0.038	<0.001	0.200[e]	0.046	<0.001	0.200[e]	0.200[e]	<0.001	0.002
One-way ANOVA[d]	*P*-value	0.058	<0.001	0.014	<0.001	<0.001	<0.001	<0.001	0.167	0.473	<0.001

[a]Means within a column with different letters differ significantly ($p < 0.05$), evaluated either using the multiple comparison Tukey's HSD or Tamhane's T2 tests, depending on the fulfilment or not of the homoscedasticity requirement, respectively.
[b]Homoscedasticity among cultivars was tested by means of the Levene test: homoscedasticity, *P*-value>0.05; heteroscedasticity, *P*-value<0.05.
[c]Normal distribution of the residuals was evaluated using Kolmogorov-Smirnov with Lilliefors correction test (n>20).
[d]*P*<0.05 meaning that the mean value of the evaluated parameter of at least one cultivar differs from the others (in this case multiple comparison tests were performed).
[e]This is a lower bound of the true significance.

Table 4.

Tocopherol Vitamers Composition (mg/100 g of fresh fruit) for Assembled PDO and Individual non-PDO cultivars. The results, analysed through 1-way ANOVA, are presented as mean±SD.[a]

		α-tocopherol	α-tocotrienol	β-tocopherol	γ-tocopherol	γ-tocotrienol	δ-tocopherol
Cultivar	PDO (n = 28)	33±11	0.2±0.1 a	0.19±0.05	2.1±0.5	0.17±0.05	0.04±0.01
	Ferraduel (n = 6)	32±11	0.1±0.1 ab	0.18±0.04	1.5±0.4	0.11±0.02	0.05±0.02
	Ferragnes (n = 6)	37±8	0.2±0.2 ab	0.24±0.05	1.4±0.4	0.24±0.05	0.04±0.01
	Ferrastar (n = 6)	38±7	0.2±0.2 ab	0.19±0.04	1.9±0.4	0.12±0.01	0.04±0.01
	Gloriette (n = 4)	27±3	0.11±0.03 ab	0.21±0.03	0.7±0.1	0.11±0.05	0.02±0.01
	Marcona (n = 4)	38±9	0.04±0.01 b	0.18±0.04	1.2±0.5	0.15±0.04	0.02±0.01
Homoscedasticity[b]	*P*-value	0.432	<0.001	0.465	0.001	<0.001	0.260
Normal distribution[c]	*P*-value	0.200[e]	0.024	0.013	0.060	<0.001	0.019
One-way ANOVA[d]	*P*-value	0.473	0.018	0.896	0.179	0.087	0.201

[a]Means within a column with different letters differ significantly ($p < 0.05$), evaluated either using the multiple comparison Tukey's HSD or Tamhane's T2 tests, depending on the fulfilment or not of the homoscedasticity requirement, respectively.
[b]Homoscedasticity among cultivars was tested by means of the Levene test: homoscedasticity, *P*-value>0.05; heteroscedasticity, *P*-value<0.05.
[c]Normal distribution of the residuals was evaluated using Kolmogorov-Smirnov with Lilliefors correction test (n>20).
[d]*P*<0.05 meaning that the mean value of the evaluated parameter of at least one cultivar differs from the others (in this case multiple comparison tests were performed).
[e]This is a lower bound of the true significance.

Table 5. Contingency Matrix Obtained Using LDA Based on TGA Profiles of Almonds Belonging to PDO *Amêndoa Douro* Cultivars and Five non-PDO Cultivars.

Actual group	Predicted group						Total	Sensitivity (%)
	PDO	Ferraduel	Ferragnes	Ferrastar	Gloriette	Marcona		
PDO	25	1	1	0	0	1	28	89
Ferraduel	0	6	0	0	0	0	6	100
Ferragnes	0	0	6	0	0	0	6	100
Ferrastar	0	0	0	6	0	0	6	100
Gloriette	0	0	0	0	4	0	4	100
Marcona	0	0	0	0	0	4	4	100
Total	25	7	7	6	4	5	54	94
Specificity (%)	100	86	86	100	100	80	92	

Figure 1.

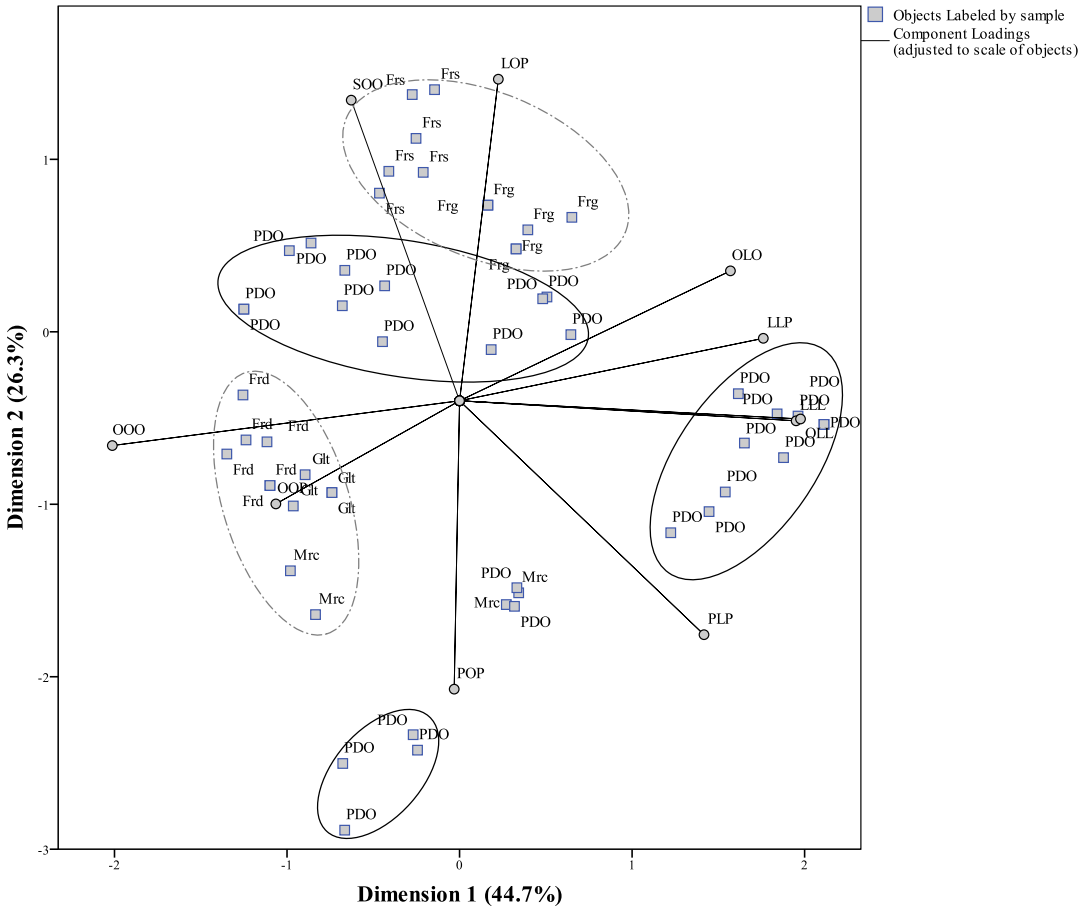
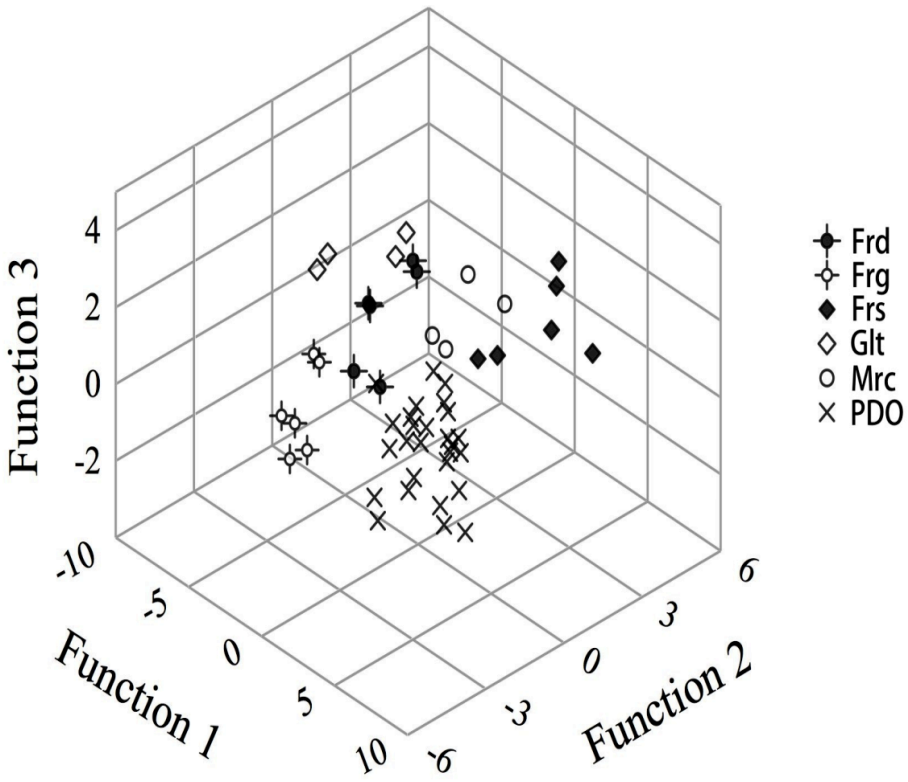
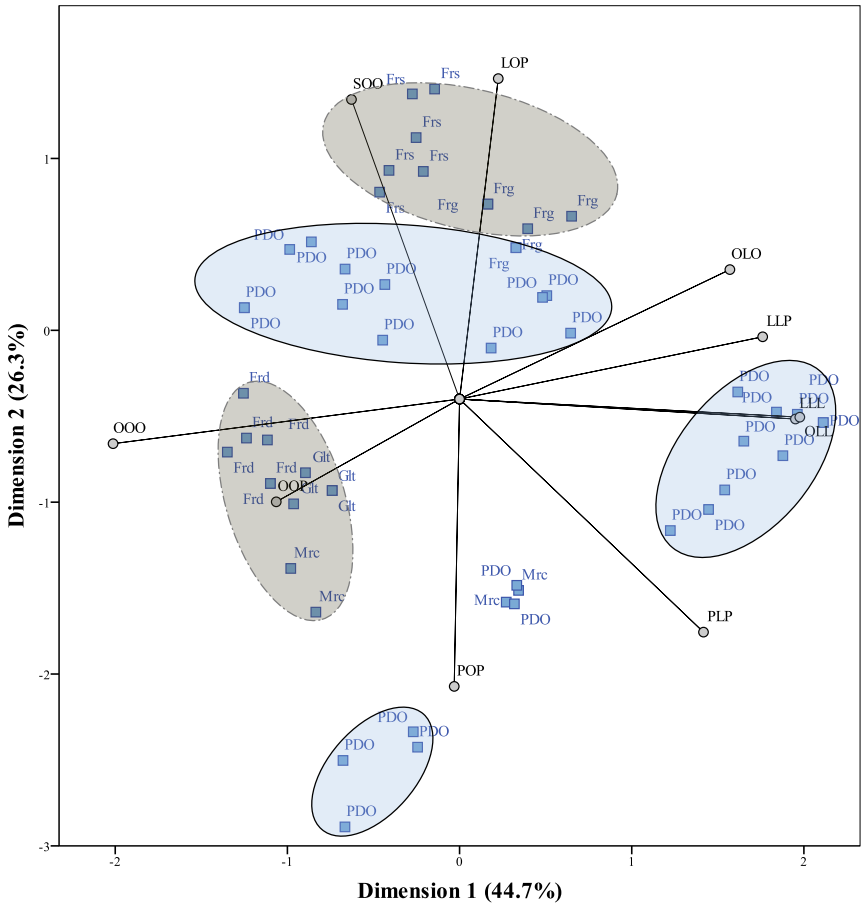
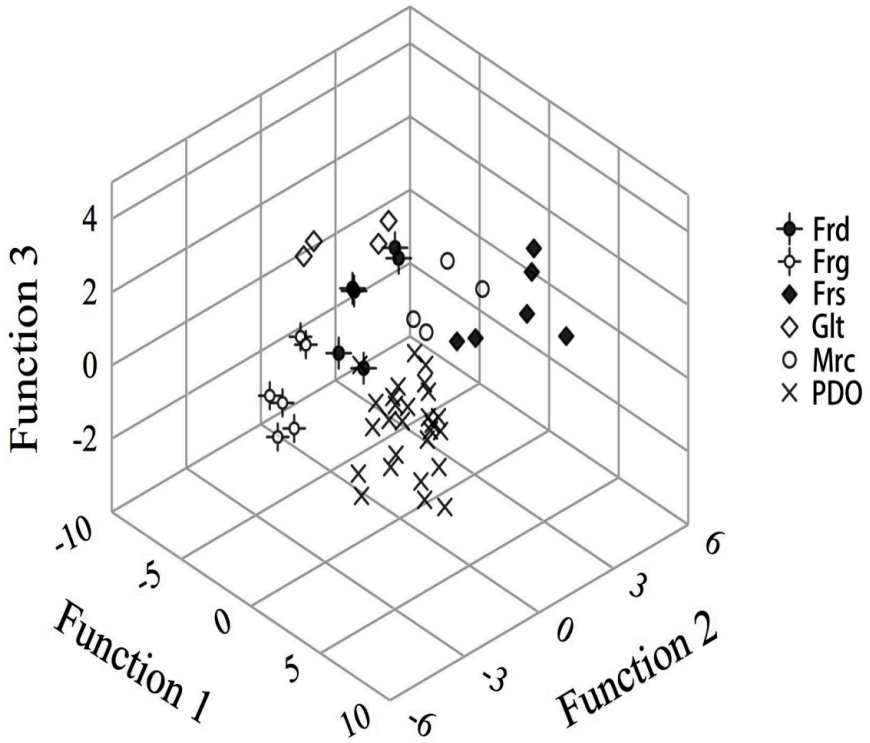

Objects Labeled by sample
Component Loadings
(adjusted to scale of objects)
Dimension 2 (26.3%)
Dimension 1 (44.7%)
LOP
SOO
OLO
LLP
OOO
PLP
POP
Frs
Frg
PDO
Frd
Glt
Mrc
OOP
OLL
LLL

Figure 2.


Function 3
4
2
0
-2
-10
-5
0
5
10
Function 1
6
3
0
-3
-6
Function 2
Frd
Frg
Frs
Glt
Mrc
PDO

TOC Graphic


LOP
SOO
OLO
LLP
OOO
PLP
POP
Frs
Frg
Frd
Glt
Mrc
PDO
OOP
OLL
Dimension 2 (26.3%)
Dimension 1 (44.7%)


Function 3
Function 1
Function 2
Frd
Frg
Frs
Glt
Mrc
PDO